

Versión “postprint” del documento publicado en:

Sánchez Fernández, Luis; Miñana Rontomé, Tello Ismael; Fernández García, Norberto; Basanta Val, Pablo; Arias Fisteus, Jesús; Congosto Martínez, María Luz; Fuentes Lorenzo, Damaris (2015) *Primeros experimentos para evaluar la calidad de fuentes de información en Twitter*. En: Asorey-Cacheda, R. et al., Actas: III Congreso Nacional de i+d en Defensa y Seguridad: DESEi+d 2015. Centro Universitario de la Defensa de Marín, 2015. Pp. 331-338.

© Rafael Asorey Cacheda, Rosa Devesa Rey, M. Mercedes Solla Carracelas, Jose M. Pousada Carballo

© De la presente edición, Centro Universitario de la Defensa de Marín, 1ª edición, 2015

# Primeros experimentos para evaluar la calidad de fuentes de información en Twitter

Sánchez Fernández, Luis <sup>1,\*</sup>, Miñana Rontomé, Tello <sup>1</sup>, Fernández García, Norberto <sup>2</sup>, Basanta Val, Pablo <sup>1</sup>, Arias Fisteus, Jesús <sup>1</sup>, Congosto, Mariluz <sup>1</sup>, Fuentes-Lorenzo, Damaris <sup>1</sup>

<sup>1</sup> Departamento de Ingeniería Telemática,  
Universidad Carlos III de Madrid  
Av. Universidad, 30, E-28911 Leganés (Madrid)  
webtlab@inv.it.uc3m.es

<sup>2</sup> Centro Universitario de la Defensa  
Escuela Naval Militar  
Plaza de España, s/n, 36920 Marín (Pontevedra)  
norberto@ cud.uvigo.es

\* Autor responsable del trabajo; luiss@it.uc3m.es

---

**Abstract:** Las redes sociales en línea gozan de una gran popularidad. A su vez, es bien conocido que buena parte de la información que se difunde en dichas redes es en muchos casos de baja calidad, rumores o simplemente información falsa. En esta comunicación se presentan los primeros resultados de un trabajo que se está realizando para evaluar la calidad de fuentes de información en línea.

**Keywords:** Seguridad; redes sociales; Twitter; calidad de fuentes de información

---

## 1. Introducción

Las redes sociales en línea [1] están suponiendo una revolución en la difusión y crecimiento de la información accesible a través de Internet adicional a la que supuso el propio desarrollo de la Web [2,3]. Al igual que pasa con la Web, la naturaleza abierta de estas redes es una de las causas de su éxito. A su vez, también es una de sus debilidades: cualquiera puede difundir información en una red social y no hay ninguna garantía de que dicha información sea verídica [4–6]. En otras ocasiones, las redes sociales son utilizadas por organizaciones y empresas como medio para intentar influir en la opinión pública [7,8] o con fines publicitarios [9]. Desafortunadamente, también pueden darse situaciones en que las redes sociales sean una herramienta en la comisión de delitos [10].

En diferentes ámbitos (búsqueda de información [11], protección contra mensajes basura (spam) [12], seguridad [10], análisis de opinión [13], salud pública [14], etc.), es interesante disponer de herramientas que puedan ayudar a evaluar la (buena o mala) calidad de fuentes de datos de redes sociales.

Esta comunicación presenta los resultados de algunos experimentos que se han realizado con datos tomados de un conjunto de cuentas Twitter, con el objetivo de avanzar en el desarrollo de herramientas que ayuden a evaluar la calidad de fuentes de información en Twitter.

El resto del artículo está organizado de la siguiente forma. La sección 2 enuncia un conjunto de hipótesis de las que se parte. La sección 3 describe la metodología que se va a utilizar para validar dichas hipótesis. La sección 4 describe los tipos de cuentas en Twitter que se han analizado así como las principales fuentes que se han utilizado para su selección. La sección 5 presenta los resultados que se han obtenido en los experimentos realizados. La siguiente sección presenta una somera discusión de dichos resultados. La comunicación termina presentando algunos trabajos relacionados y detallando las conclusiones obtenidas en este estudio, así como posibles líneas futuras de continuación del mismo.

## **2. Hipótesis de partida**

El primer paso en el trabajo que se ha realizado es plantear algunas hipótesis cuya validez pueda permitir identificar fuentes de información de calidad. Las hipótesis que vamos a explorar parten de la idea de que las fuentes de información de calidad (medios de comunicación, investigadores, periodistas, escritores, etc.) utilizan un lenguaje más culto que el de los usuarios normales y las fuentes de información poco fiables. Vamos a explorar esta idea desde dos puntos de vista. Por una parte vamos a evaluar la complejidad de la estructura sintáctica de las sentencias escritas en una cuenta de Twitter. Por otra parte, vamos a medir el registro lingüístico de cada cuenta (es decir, el número de palabras distintas utilizadas).

En resumen, planteamos las siguientes dos hipótesis:

**H1** Las fuentes de información de calidad utilizan estructuras sintácticas más complejas que los usuarios normales y los difusores de información falsa.

**H2** Las fuentes de información de calidad utilizan registros lingüísticos más amplios.

## **3. Metodología utilizada**

Es posible definir multitud de indicadores relacionados con las dos hipótesis presentadas en la sección anterior. En este primer trabajo hemos analizado solamente tres.

Los primeros dos indicadores están relacionados con la primera hipótesis y tratan de medir la complejidad sintáctica.

1. Indicador 1: porcentaje de tuits de una cuenta que contienen oraciones subordinadas.

2. Indicador 2: número medio de sintagmas adjetivales que contiene cada tuit de la cuenta.

Para medir estos dos indicadores se ha utilizado un analizador sintáctico para procesamiento de lenguaje natural desarrollado en la Universidad de Stanford [15].

Para medir la calidad del registro lingüístico se ha utilizado la siguiente estrategia. Se ha utilizado una lista de frecuencias del uso de las palabras tomada del Wiktionary [16]. Esta lista cuenta la frecuencia de

aparición de las palabras más frecuentes en todos los libros que aparecen en el proyecto Gutenberg [17] (más de 35,000 palabras). El indicador 3 mide el porcentaje de las 20,000 palabras menos frecuentes utilizadas en esta lista que han sido utilizadas en alguno de los últimos 1,000 tuits de la cuenta estudiada.

Para evaluar la validez de las hipótesis de partida se han medido estos indicadores para un conjunto de cuentas de Twitter suficientemente grande como para que los resultados sean estadísticamente significativos clasificados en diferentes tipos. A continuación los valores de los indicadores obtenidos para cada tipo de cuenta se han presentado en forma de box-plot, tal y como se describe en la sección de resultados.

#### **4. Recolección de datos**

Para realizar los experimentos se seleccionaron cuentas en Twitter de los siguientes tipos:

- Cuentas de medios de comunicación, todas ellas verificadas oficialmente por Twitter y obtenidas de [18].
- Cuentas de investigadores, profesores de universidad y periodistas, obtenidas de los seguidores de cuentas de universidades prestigiosas del mundo como Stanford y Cambridge.
- Cuentas de personajes famosos (políticos, actores, deportistas, etc.) y verificadas.
- Cuentas “fake”: cuentas que parodian a los personajes de la categoría anterior, la mayoría de ellas obtenidas de [19–21].
- Cuentas bot (cuyos tuits son producidos por programas de ordenador), obtenidas de [22] y otras por medio de búsqueda manual.
- Cuentas aleatorias. La idea de esta categoría es disponer de cuentas de usuarios normales que no pertenezcan a ninguna de las categorías anteriores. Para obtener las cuentas aleatorias nos conectamos al `Stream API` de Twitter [23] y escogimos las primeras 40 cuentas en inglés y que hayan tuiteado al menos 10 veces desde que se creó la cuenta para tener algunos datos con los que poder realizar el análisis.

Para cada tipo de cuenta de Twitter se han analizado 40 cuentas. Esto debe ser suficiente, por ejemplo, para entrenar un clasificador con la técnica clásica del análisis discriminante de Fisher [24] utilizando muestras de 20 dimensiones [25].

Como tenemos 6 tipos de cuentas y 40 cuentas de cada tipo, en total se han analizado 240 cuentas.

Debido a que las herramientas que se han utilizado en este estudio están diseñadas para procesar texto en inglés, solamente se han considerado cuentas de Twitter que tuitean principalmente en inglés.

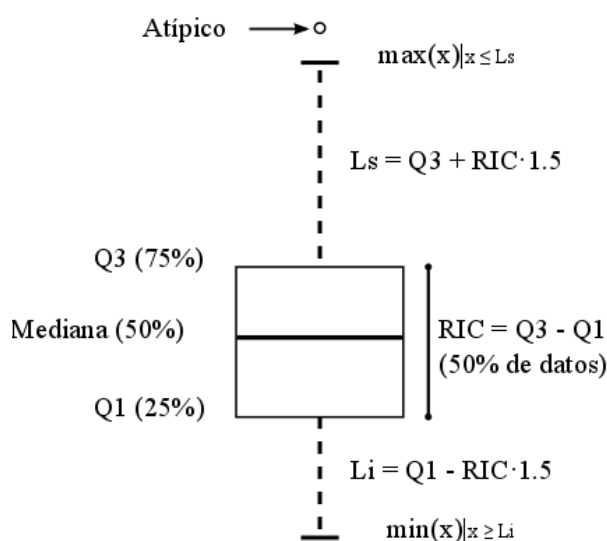
#### **5. Resultados obtenidos**

Cada uno de los tres indicadores descritos en la sección 3 se ha calculado para cada una de las 240 cuentas consideradas. Para visualizar los resultados obtenidos estos se presentan por medio de un diagrama de cajas o box-plot [24].

La figura 1<sup>1</sup> representa un diagrama de caja. Un diagrama de caja es un gráfico, basado en cuartiles, mediante el cual se visualiza un conjunto de datos. Está compuesto por un rectángulo, la "caja", y dos brazos, los "bigotes". La información que suministra el diagrama de caja, como se puede ver en la figura 1 es la siguiente:

- Primer cuartil (Q1)
- Segundo cuartil o mediana. Es la raya dentro de la caja.
- Tercer cuartil (Q3)
- Los bigotes, las líneas que se extienden desde la caja, se extienden hasta los valores máximo y mínimo de la serie o hasta 1.5 veces el Rango Inter Cuartílico (Q3 - Q1).
- Cuando los datos se extienden más allá de esto, significa que hay valores atípicos en la serie, que se marcan con pequeños círculos.

**Figure 1.** Diagrama de caja o box-plot



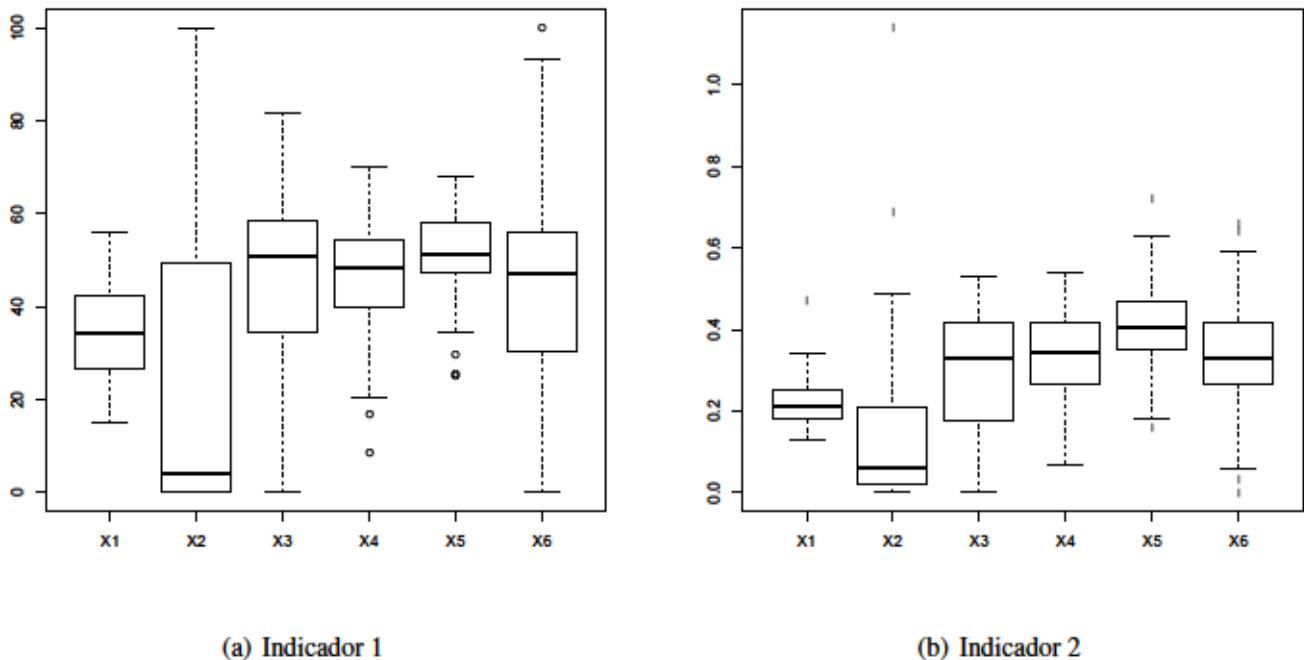
Los resultados obtenidos para cada uno de los indicadores se presentan en las figuras 2 y 3. En cada una de estas figuras, el significado de las etiquetas en el eje de abscisas es el siguiente: X1: cuentas de medios de comunicación; X2: cuentas bot; X3: cuentas fake; X4: cuentas de famosos; X5: cuentas de investigadores, profesores y periodistas; X6: cuentas aleatorias.

## 6. Discusión

El trabajo presentado en esta comunicación debe servir para identificar cuáles de los 3 indicadores considerados deben de formar parte del conjunto de indicadores adecuados para detectar fuentes de información de calidad y cuáles no.

<sup>1</sup>“Boxplot” de Jumanbar - Trabajo propio. Disponible bajo la licencia CC BY-SA 3.0 vía Wikimedia Commons - <http://commons.wikimedia.org/wiki/File:Boxplot.svg#/media/File:Boxplot.svg>

Figure 2. Resultados para los indicadores sintácticos



Los resultados obtenidos para el indicador 1 no permiten distinguir las cuentas de unos tipos y otros, por lo que este indicador no debe ser incluido en el conjunto de indicadores para detectar fuentes de información de calidad.

Los resultados para el indicador 2 se pueden calificar de moderados, porque se observan valores mayores en la categoría de profesores, investigadores y periodistas frente a las demás, aunque las diferencias no sean demasiado significativas.

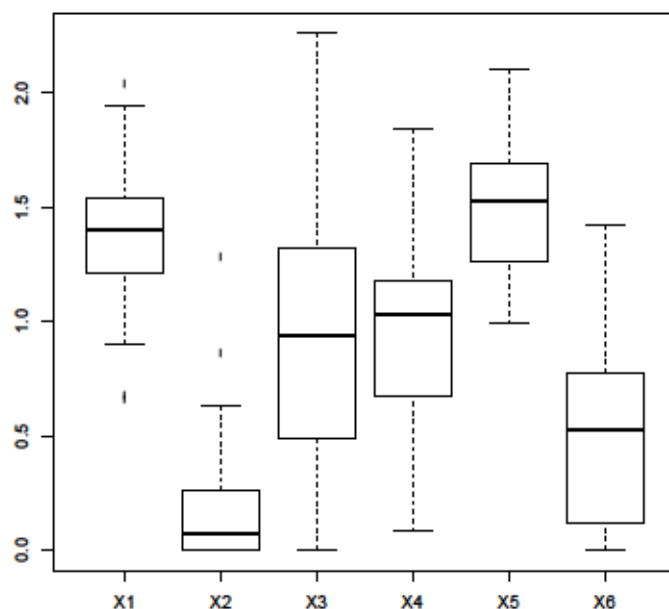
Finalmente, los resultados para el indicador 3 distinguen de forma clara a las cuentas de medios de comunicación y las de profesores, investigadores y periodistas. Consideramos ambos tipos de cuentas como cuentas de calidad y por lo tanto el indicador 3 también se considera válido.

El hecho de que el indicador 2 no distinga los medios de comunicación de los demás tipos de cuentas debería ser tenido en cuenta. Una posibilidad es procesar las cuentas en dos etapas, de manera que en una primera etapa se identifiquen los medios de comunicación (y no se utilice el indicador 2) y a las cuentas restantes se les aplique un segundo procesamiento en el que se incluirían (entre otros) los indicadores 2 y 3.

## 7. Trabajos relacionados

Existen varios trabajos relacionados con el estudio de la fiabilidad de la información difundida en redes sociales. En [26] se analiza la fiabilidad de los tuits difundidos durante el desastre de Fukushima. Este estudio se centra en los denominados tuits sintetizados, en los que un usuario sintetiza información proveniente de cierta fuente de información (en contraposición a un retuiteo puro) y que utilizaban el hashtag #fukushima. Se encontró que el 70% de los tuits provenían de fuentes de información fiables, con un porcentaje más bajo cuando los tuits provenían de cuentas con un perfil anónimo.

**Figure 3.** Resultados para el indicador 3: Porcentaje de las 20,000 palabras menos frecuentes utilizadas



Otro trabajo en la misma línea es el de [4]. En este trabajo se analizó la difusión de tuits en el terremoto de Chile de 2010. En este estudio se concluyó que los rumores en Twitter tienden a ser más cuestionados que las noticias verídicas, lo que puede ser un instrumento en si mismo para su detección.

La detección de spammers puede considerarse como un apartado dentro del análisis de la fiabilidad de fuentes de información que ha sido estudiado en numerosos trabajos [12,27].

## 8. Conclusiones y Trabajo Futuro

En el estudio que presentamos en este trabajo se han evaluado tres indicadores con el objetivo de identificar fuentes de información de calidad de Twitter. De los tres indicadores considerados, uno no ha proporcionado los resultados esperados mientras que los resultados obtenidos para los otros dos son prometedores (mejores en el caso del indicador 3 que en el del 2).

Como se ha explicado en la introducción, el estudio que se presenta en este trabajo es un primer paso que debe ser continuado en trabajos posteriores. A continuación se esbozan varias líneas de trabajo que planeamos desarrollar en el futuro. En primer lugar, es necesario ampliar el conjunto de indicadores que se utilizan. Dentro de este apartado se deben ampliar los indicadores sintácticos y del registro del lenguaje. Asimismo, se podrían considerar otros tipos de indicadores, como ejemplo estudiar los patrones de comportamiento temporal de los tuits de una cuenta.

Una vez que se disponga de un conjunto de indicadores adecuado, será necesario entrenar un clasificador utilizando técnicas de aprendizaje máquina.



Finalmente, será necesario realizar una evaluación del clasificador así obtenido utilizando datos de un volumen de cuentas elevado (cientos de miles) para poder garantizar que los resultados obtenidos sean fiables.

La explotación en un entorno realista de una herramienta como la que se pretende desarrollar involucraría el procesado de grandes volúmenes de datos (Big Data) para aplicar estas técnicas a un gran número de cuentas de Twitter. Para tratar con esto sería interesante utilizar modelos predecibles de procesado distribuido de flujos, como el propuesto en [28].

## Agradecimientos

Este trabajo está parcialmente financiado por el Ministerio de Economía y Competitividad a través del proyecto “HERMES-SMARTDRIVER” (TIN2013-46801-C4- 2-R) y por la Comunidad de Madrid a través del proyecto “eMadrid” (S2013/ICE-2715).

## Referencias

1. N. B. Ellison *et al.*, “Social network sites: Definition, history, and scholarship,” *Journal of Computer-Mediated Communication*, vol. 13, no. 1, pp. 210–230, 2007.
2. T. Berners-Lee, M. Fischetti, and M. L. Foreword By-Dertouzos, *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. Harper Information, 2000.
3. B. A. Huberman and L. A. Adamic, “Internet: growth dynamics of the world-wide web,” *Nature*, vol. 401, no. 6749, pp. 131–131, 1999.
4. M. Mendoza, B. Poblete, and C. Castillo, “Twitter under crisis: Can we trust what we rt?” in *first workshop on social media analytics*. ACM, 2010, pp. 71–79.
5. D. Mocanu, L. Rossi, Q. Zhang, M. Karsai, and W. Quattrociocchi, “Collective attention in the age of (mis)information,” *Computers in Human Behavior*, no. 0, pp. –, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0747563215000382>
6. A. Bessi, M. Coletto, G. A. Davidescu, A. Scala, G. Caldarelli, and W. Quattrociocchi, “Science vs conspiracy: collective narratives in the age of (mis) information,” *PLoS ONE*, vol. 10, no. 2, 2015.
7. M. L. Congosto, “Elecciones europeas 2014: Viralidad de los mensajes en twitter,” *Redes: revista hispana para el análisis de redes sociales*, vol. 26, no. 1, pp. 23–52, 2015.
8. M. L. Congosto, “Viralidad de los mensajes en twitter en las campañas electorales,” in *III Congreso Internacional en Comunicación Política y Estrategias de Campaña*, 2014.
9. J. A. Chevalier and D. Mayzlin, “The effect of word of mouth on sales: Online book reviews,” *Journal of marketing research*, vol. 43, no. 3, pp. 345–354, 2006.
10. C. C. Yang and T. D. Ng, “Terrorism and crime related weblog social network: Link, content analysis and information visualization,” in *Intelligence and Security Informatics, 2007 IEEE*. IEEE, 2007, pp. 55–58.
11. S. P. Borgatti and R. Cross, “A relational view of information seeking and learning in social networks,” *Management science*, vol. 49, no. 4, pp. 432–445, 2003.
12. D. Gayo-Avello, “Nepotistic relationships in twitter and their impact on rank prestige algorithms,” *Information Processing & Management*, vol. 49, no. 6, pp. 1250–1280, 2013.



13. F. Li and T. C. Du, "Who is talking? an ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs," *Decision Support Systems*, vol. 51, no. 1, pp. 190–197, 2011.
14. S. A. Moorhead, D. E. Hazlett, L. Harrison, J. K. Carroll, A. Irwin, and C. Hoving, "A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication," *Journal of medical Internet research*, vol. 15, no. 4, 2013.
15. D. Klein and C. D. Manning, "Accurate unlexicalized parsing," in *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*. Association for Computational Linguistics, 2003, pp. 423–430.
16. "Wiktionary:frequency lists," [http://en.wiktionary.org/wiki/Wiktionary:Frequency\\_lists#English](http://en.wiktionary.org/wiki/Wiktionary:Frequency_lists#English) (visitado el 12 de mayo de 2015), 2015.
17. "Free ebooks - project gutenber," [https://www.gutenberg.org/wiki/Main\\_Page](https://www.gutenberg.org/wiki/Main_Page) (visitado el 12 de mayo de 2015), 2015.
18. N. Bremmen, "The 100 most influential news media twitter accounts," <http://memeburn.com/2010/09/the-100-most-influential-news-media-twitter-accounts/> (visitado el 7 de mayo de 2015), 2010.
19. C. McCann, "The list: Ten fake twitter accounts," <http://www.ft.com/cms/s/2/c843804c-3b21-11e2-b3f0-00144feabdc0.html#axzz2jz08haZW> (visitado el 7 de mayo de 2015).
20. I. Paul, "15 fake and funny twitter accounts," [http://www.pcworld.com/article/159492/fake\\_funny\\_twitter.html](http://www.pcworld.com/article/159492/fake_funny_twitter.html) (visitado el 7 de mayo de 2015), 2009.
21. E. Levine, "The 15 best fake twitter accounts of 2012," <http://heavy.com/comedy/2012/12/the-15-best-fake-twitter-accounts-of-2012/> (visitado el 7 de mayo de 2015), 2012.
22. K. Knibbs, "The 8 best twitter bots you aren't following," <http://www.digitaltrends.com/social-media/the-10-best-twitter-bots-you-arent-following/> (visitado el 7 de mayo de 2015), 2013.
23. "The streaming apis," <https://dev.twitter.com/streaming/overview> (visitado el 7 de mayo de 2015), 2015.
24. D. Peña, *Análisis de datos multivariantes*. Mc Graw Hill Interamericana de España, S.A.U., 2002.
25. S. J. Raudys and A. K. Jain, "Small sample size effects in statistical pattern recognition: Recommendations for practitioners," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 13, no. 3, pp. 252–264, 1991.
26. R. Thomson, N. Ito, H. Suda, F. Lin, Y. Liu, R. Hayasaka, R. Isochi, and Z. Wang, "Trusting tweets: The fukushima disaster and information source credibility on twitter," in *9th International ISCRAM Conference*, 2012, pp. 1–10.
27. G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in *26th Annual Computer Security Applications Conference*. ACM, 2010, pp. 1–9.
28. P. Basanta, N. Fernandez-Garcia, A. Wellings, and N. Audsley, "Improving the predictability of distributed stream processors," *Future Generation Computing Systems*, 2015.